تحديد العوامل المؤثرة في وفيات وباء كورونا باستعمال انموذج الاستجابة الثنائية

الاستاذ المساعد الدكتور مهدي وهاب نعمة الباحث علي محمد جواد

كلية الادارة والاقتصاد / قسم الاحصاء

جامعة كربلاء

المدرس المساعد علي عبد الزهرة حسن

جامعة البصرة / كلية الادارة والاقتصاد / قسم الاحصاء

المستخلص :

يعد وباء كورونا من الاسباب الرئيسة لوفيات الكثير من السكان في الوقت الحالي وللتعرف على أكثر هذه العوامل تأثيرا في وفيات المصابين بوباء كورونا تم استعمال انموذج الاستجابة الثنائية او الانحدار اللوجستي الثنائي. وقد تم اختيار ثمانية متغيرات لتحديد تأثيرها في وفيات المرضى المصابين بوباء كورونا باستعمال طريقتين للاختيار، هما طريقة التقدم الامامي وطريقة الحذف العكسي. واظهرت النتائج على وفق طريقة التقدم الامامي والحذف العكسي أن اهم العوامل هي (عمر المريض، التدخين). وقد تم اختبار جودة التوفيق للنماذج النهائية الناتجة من طريقتي الاختيار باستعمال اختبارين هما اختبار هوزمر – ليمشو (H&L)، واختبار الانحراف (D) وبين كليهما ان الانموذجين المقدرين معنويين، كما تضمنت الدراسة مقارنة بين الانموذجين باستعمال معيارين للمقارنة وهما معيار نسبة الامكان الاعظم (MLR)، ومعيار جدول التصنيف (CTC) واثبت كل من المعيارين ان الانموذج الناتج من الطريقتين هو الانموذج الانسب.

Determine the factors affecting the deaths of the Corona epidemic using the Binary response model

Professor Dr. Mahdi Wahhab Nehmeh **Researcher Ali Muhammad Jawad**

Faculty of Administration and Economics / Department of Statistics

University Karbala

Assistant teacher Ali Abdel-Zahra Hassan

College of Administration and Economics / Statistics Department

University Basra

Abstract :

The Corona epidemic is one of the main causes of death at the present time. To identify the most influential factors in the death of those who were infected with the Corona epidemic,a Binary response model or a binary logistic regression was used eight variables were chosen to determine their effect on the mortality of patients with the Corona epidemic by using two selection methods, the forward advance method and the reverse deletion method, respectively. The results show, according to the method of forward progression and reverse deletion, that the most important factors are (patient age, smoking). The quality of fitment was tested for the final models resulting from the two selection methods by using two tests, namely the Hosmer - Limchu test (H&L), and the deviation test (D). Both of them showed that the two estimated models are significant, and the study also included a comparison between the two models using two criterias for comparison, which are the greatest likelihood ratio (MLR) and the Classification Table Criteria (CTC). Both criterias proved that the resulting model from the two methods is the most appropriate one

.

المقدمة :

يشكل وباء كورونا سبب للوفيات في كثير من دول العالم خصوصاً في الدول التي انتشر فيها الوباء بشكل اوسع، وهي المسؤولة عن نسبة كبيرة من إجمالي الوفيات في هذه الدول. وهذا ما ينطبق على واقعنا في العراق، إذ كثر الحديث في عام 2020 عن هذه المرض المنتشر جداً، و يعد انموذج الانحدار اللوجستي الثنائي من النماذج اللاخطية الملائمة لهذه الظواهر إذ يصف العلاقة بين متغير تابع ثنائي القيمة اي يأخذ قيمتين هما الصفر لاحتمال عدم حدوث حدث معين و الواحد الصحيح لاحتمال حدوث ذلك الحدث و المتغيرات المستقلة تأخذ قيما وصفية او كمية , ويمكن تحويل انموذج الانحدار اللوجستي الى الصيغة الخطية وذلك باستعمال تحويل اللوجت (logit transformation), كما تعد طريقة الامكان الاعظم من اهم الطرائق المستعملة لتقدير معالمه . ولزيادة كفاءة الانموذج وخصوصا عندما يكون عدد المتغيرات المستقلة كبيرا يلجأ الباحثون الى استعمال طرائق اختيار المتغيرات المستقلة مثل طريقة التقدم الامامي وطريقة الحذف العكسي.

اهمية البحث

تكمن اهمية هذا البحث في تحديد العوامل التي تؤثر في وفيات وباء كورونا كون دراسة جميع العوامل او المتغيرات تحتاج الى جهد ووقت ومال اضافة الى ذلك قد تكون النتائج غير دقيقة.

هدف البحث

يهدف البحث الى تحديد أفضل أنموذج انحدار لوجستي لوفيات وباء كورونا بأقل عدد ممكن من المتغيرات المستقلة وذلك باستعمال طريقتين لاختيار المتغيرات المستقلة وهما طريقتا التقدم الامامي والحذف العكسي. والمقارنة بين نتائجهما عن طريق استعمال بعض معايير المفاضلة بين النماذج.

مشكلة البحث :

نظراً لأهمية وباء كورونا وتأثيره في حياة الناس لكونه من الامراض المستجدة والواسعة الانتشار في المجتمع فتكون سبباً لوفاة الكثير، لذا سيتم تطبيق انموذج الانحدار اللوجستي على المرضى الراقدين في مستشفى سوق الشيوخ لسنة 2020 لمعرفة العوامل التي تؤثر في وفيات هؤلاء المصابين ، اذ ان هناك العديد من العوامل بعضها له تأثير بدرجة قوية وبعضها بدرجة اقل إذ يمكن ان يستبعد من الانموذج لأسباب عديدة منها ان تأثيره يمكن ان يكون مماثلا لتأثير عامل اخر في الانموذج.

الجانب النظري

مفهوم الانحدار اللوجستي

عند دراسة سلوك العديد من الظواهر الطبيعية نجدها تسلك سلوكاً غير خطي، ولكي نحلل تلك الظواهر فأننا نستعمل النماذج اللاخطية، كما ان الاستجابة في بعض الظواهر تأخذ قيمتين فقط (binary) او قيم تعبر عن الاوجه المختلفة للظاهرة محل الدراسة وهذه القيم ليس لها معنى قائم بحد ذاته انما هي مجرد دلالة على وجود الحدث من عدمه. ان الانموذج الملائم لهذا النوع من البيانات هو انموذج الانحدار اللوجستي والذي يمكن تعريفه بأنه اسلوب احصائي لوصف العلاقة بين المتغير التابع الوصفي ومتغير واحد او اكثر من متغير مستقل وصفي او كمي [7] وسنختصر في دراستنا على الظواهر التي يكون فيها المتغير التابع يأخذ قيمتين فقط وهو ما يسمى بأنموذج الانحدار اللوجستي ذي الاستجابة الثنائية إذ يفترض هذا الانموذج ما يأتي [5] :

فرضيات البحث:

1- المتغير التابع متغير وصفي يأخذ قيمتين فقط هما (الصفر والواحد) اما المتغيرات المستقلة فقد تكون كمية او وصفية.
2- وجود علاقة خطية بين لوغاريتم نسبة الرجحان (Odd ratio) والمتغيرات المستقلة.
3- عدم وجود علاقة تامة أو شبه تامة بين المتغيرات المستقلة بعضها مع بعض، وذلك حتى نضمن الحصول على تقديرات غير متحيزة وأخطاء معيارية صغيرة. فضلاً عن العينة يجب ان تكون عشوائية والبيانات خالية من الأخطاء والقيم المتطرفة.

يعد الفرض الاول هو الفرض الاساسي الذي يبنى عليه انموذج الانحدار اللوجستي إذ يرمز لوقوع الحدث بالرمز (1) وباحتمال (P) وعدم وقوعه بالرمز (0) وباحتمال (1-P) [4] و حدوثهما من عدمه يعد دالة للمتغير او المتغيرات المستقلة . و بذلك فالمتغير التابع يتبع توزيعاً احتماليا متقطعا وهو توزيع برنولي ويمكن كتابته بداله على وفق الانموذج اللوجستي (logistic model) بالشكل الآتي [6] :

$$P = \frac{e^{XB}}{1 + e^{XB}} \qquad (1)$$

$$1 - P = 1 - \frac{e^{XB}}{1 + e^{XB}}$$

$$q = \frac{1}{1 + e^{XB}} \qquad (2)$$

اذ $q = 1 - p$

m.....j=1,2,3 (m) هي عدد المتغيرات المستقلة

X: متجه يمثل المتغيرات المستقلة عددها (..... j=1,2

β= متجه معالم الانموذج وتكون مجهولة القيم يتم تقديرها من الانموذج

ولتحويل انموذج الانحدار اللوجستي الى الصيغة الخطية تواجهنا مشكلة وهي ان قيمة المتغير التابع هي بين (الصفر والواحد) ومن المعلوم ان قيم المتغير التابع في الانحدار الخطي تقع بين $(-\infty, \infty)$ وللتغلب على هذه المشكلة يستعمل التحويل الرياضي الآتي $(\frac{p}{1-p})$ وهو ما يعرف بنسبة الرجحان وحدوده بين $(0, \infty)$, نلاحظ ان هذا التحويل قد حل مشكلة الحد الاعلى اي انه حل نصف المشكلة [8] وبقى النصف الاخر المتمثل بالحدود الدنيا لذلك نأخذ اللوغاريتم الطبيعي للمقدار $(\frac{p}{1-p})$ وبهذا يكون مجاله بين $(-\infty, \infty)$ ومن ثم توصلنا الى حدود المتغير التابع للانحدار الخطي [3] و كما مبين بالاتي :

$$\frac{\mathrm{p}}{1-\mathrm{p}} = \frac{\frac{e^{XB}}{1+e^{XB}}}{\frac{1}{1+e^{XB}}} \quad \rightarrow \quad \frac{\mathrm{p}}{1-\mathrm{p}} = e^{XB}$$

وبأخذ اللوغاريتم الطبيعي للطرفين

$$\ln(\frac{\mathrm{p}}{1-\mathrm{p}}) = \ln(e^{XB})$$

$$\ln(\frac{\mathrm{p}}{1-\mathrm{p}}) = XB \qquad (3)$$

ويسمى المقدار $\ln(\frac{p}{1-p})$ باللوغاريتم الطبيعي لنسبة الرجحان أو قد يسمى باللوجت (logit) وتعد طريقة الامكان الاعظم (Maximum Likelihood method) واحده من الطرائق المستعملة لتقدير المعالم (β_j) في مثل هذه النماذج الرياضية [7]

اختبار جودة التوفيق لأنموذج الانحدار اللوجستي

اختبار مربع كاي لنسبة الترجيح (الانحراف D)

يعتمد هذا الاختبار على مبدأ هو المقارنة بين القيم المشاهدة والقيم المتوقعة للأنموذج الذي يحتوي الحد الثابت فقط (constant) والأنموذج الذي يحتوي المتغيرات كافة إذ يعتمد في ترجيحه على الفرق بين نسبة الامكان الاعظم (MLR) للأنموذج الذي يحتوي الحد الثابت فقط ونسبة الامكان الاعظم للأنموذج الذي يحتوي على جميع المتغيرات [6] وحسب الفرضية الاحصائية الآتية:

(4)

H_0 : الانموذج المقدر غير ملائم

H_1: الانموذج المقدر ملائم

ويمكن ان يحسب حسب الصيغة الآتية[6]:

$$D = MLR_{(\text{للانموذج بالحد الثابت فقط})} - MLR_{(\text{للانموذج مع المتغيرات})} \qquad (5)$$

D :الانحراف لنسبة الرجحان ويتبع توزيع مربع كاي بدرجة حرية تساوي عدد المتغيرات المستقلة في الانموذج اذا كانت المتغيرات كمية وعدد اوجه المتغيرات المستقلة اذا كانت وصفية

$MLR_{(\text{للانموذج بالحد الثابت فقط})}$: نسبة الامكان الاعظم للأنموذج الذي يحتوي الحد الثابت فقط .

$MLR_{(\text{للانموذج مع المتغيرات})}$: نسبة الامكان الاعظم للأنموذج الذي يحتوي على جميع المتغيرات

و إذا كانت قيمة (p-value) لـ (D), اقل من مستوى المعنوية المحدد (α) نرفض فرضية العدم ويكون الانموذج المقدر معنويا.

اختبار هوزمر – ليمشو Hossmer- Lemshow tast (H&L)

هو احد اهم اختبارات جودة التوفيق لأنموذج الانحدار اللوجستي يعتمد على مدى قرب الاحتمالات المشاهدة والاحتمالات المتوقعة ويبنى على أساس تقسيم الحالات المدروسة الى عشر مجموعات على شكل أعمدة اما الصفوف فتقسم على اساس القيم المشاهدة للمتغير المعتمد وهما الصفر والواحد[5] وهو يختبر الفرضية الاحصائية الآتية :

H_0 : الانموذج المقدر يوافق البيانات بشكل جيد (6)

H_1 : الانموذج المقدر لا يوافق البيانات بشكل جيد

ويمكن ان يحسب من الصيغة الاتية :

$$x^2 = \sum_{k=1}^{10} \frac{(O_{1k} - n_k \bar{p}_k)^2}{n_k \bar{p}_k (1 - \bar{p}_k)} \qquad (7)$$

O_{1k}: التكرار المشاهد للمجموعة (k) *عند الاستجابة* y=1

$\bar{p}_k$: متوسط الاحتمالات المتوقعه للمجموعة (k)

n_k : مجموع التكرار المشاهد في المجموعة (k) K=1,2,.... 10

وتتم مقارنة قيمة (p-value) لـ (x^2) بدرجة حرية (k-2) مع مستوى المعنوية المحدد(α) فأذا كانت قيمة (p-value) اقل من مستوى المعنوية فيتم رفض فرضية العدم اما اذا كانت قيمة (p-value) اكبر من مستوى المعنوية فلا نرفض فرضية العدم وهذا يعني ان الانموذج يوافق البيانات بشكل جيد .

اختبار الدلالة الاحصائية لكل متغير

بعد ان تم التعرف على الطرائق التي يتم فيها فحص ملاءمة الانموذج بشكل عام فأننا سنتعرف على الطرائق التي يمكن عن طريقها فحص كل متغير مستقل على حده وهل هذا المتغير له دلالة احصائية في الانموذج المقدر ام لا؟ ولأجراء هذا الاختبار نستعمل أحصاءة والد (Wald) التي *تختبر* الفرضية الإحصائية القائلة :

$H_O : \beta_j = 0$ j= 1,2, 3 ..... m

$H_1 : \beta_i \neq 0$ $m =$ عدد المتغيرات المستقلة (8)

ان احصاءه والد (Wald) يمكن ان تحسب على وفق الصيغة الآتية [2] :

$$w = \left(\frac{\hat{\beta}_j}{\widehat{SE}(\beta_j)}\right)^2 \quad (9)$$

$\hat{\beta}_j$: المعلمة المقدره للمتغير (X_j) المطلوب اختباره

$\widehat{SE}(\hat{\beta}_j)$: الخطأ المعياري للمعلمة المقدره $(\hat{\beta}_j)$

تتم مقارنة قيمة (p-value) لـ (w) بدرجة حرية (n-j)مع مستوى المعنوية المحدد (α) فأذا كانت قيمة (p-value) اقل من مستوى المعنوية المحدد فنرفض فرضية العدم، ورفض فرضية العدم يعني ان المتغير معنوي احصائياً وذو اهمية بالنسبة للنموذج. اما لاختيار المتغيرات ذات التأثير الكبير والمعنوي في الانموذج فهناك العديد من الطرائق تم اختيار طريقتين منها كما في المبحث الآتي :

طرائق اختيار انموذج الانحدار اللوجستي

طريقة التقدم الامامي (Forward Selection Method)

تفترض هذه الطريقة في البداية ان الانموذج خال من المتغيرات المستقلة اي يحتوي على الحد الثابت (Constant) فقط ثم نبدأ بإدخال المتغيرات بشكل تدريجي حسب الخطوات الآتية:

الخطوة الاولى: تكون بإضافة المتغيرات المستقلة الواحد تلو الاخر الى الانموذج الذي يحتوي على الحد الثابت فقط ومن ثم سوف تكون لدينا (j) من النماذج والذي هو بعدد المتغيرات المستقلة، ثم نحسب نسبة الامكان الاعظم لكل انموذج وحسب المتغير المضاف (MLR_j) ثم نحسب الفرق بين لوغاريتم نسبة الامكان الاعظم للأنموذج الذي يحتوي على الحد الثابت فقط و الانموذج الذي يحتوي على الحد الثابت مع متغير المضاف وحسب الصيغة الآتية [6] :

$$G_j^1 = MLR_0 - MLR_j \quad (10)$$

$$j= 1,2,.....m$$

إذ أن

MLR_0: لوغاريتم نسبة الامكان للأنموذج الذي يحتوي الحد الثابت فقط.

MLR_j: لوغاريتم نسبة الامكان للأنموذج الذي يحتوي الحد الثابت مع المتغير j في الخطوة (1).

G_j^1 : الفرق بين لوغاريتم نسبة الامكان الاعظم للأنموذج مع الحد الثابت والانموذج الذي يحتوي على الحد الثابت مع المتغير (j) و التي تتبع توزيع مربع كاي بدرجة حرية تساوي واحداً اذا كان المتغير كمياً وعدد صفات المتغير المضاف مطروحاً منها واحد اذا كان المتغير المضاف وصفياً (فئوياً) ثم نجد (p-value) لكل قيم (G_j^1) لكل متغير ويتم ترشيح المتغير الذي يحقق الشرط الآتي [1]:

$$p_j^{(1)} = \min[p_j^{(1)}] \quad (11)$$

إذ أن

$p_j^{(1)}$: قيمة (p-value) للفرق (G_j^1) للمتغير (j) في الخطوة (1) ويجب ان تكون اقل من مستوى المعنوية المحدد (α).

ثم نختبر الانموذج الذي يتضمن المتغير المضاف مع الحد الثابت بشكل عام حسب اختبار الانحراف (D) الذي تم ذكره سابقا ويحذف المتغير المضاف إذا كان الانموذج غير معنوي ويتم التوقف عن اضافة المتغيرات ويكون الانموذج الملائم هو الانموذج الذي يحتوي على الحد الثابت فقط اما إذا كان الانموذج معنوياً فيتم تثبيت المتغير المضاف ثم ننتقل الى الخطوة الآتية.

الخطوة الثانية: من الخطوة السابقة أصبح لدينا انموذج انحدار لوجستي يحتوي على الحد الثابت مع متغير مستقل واحد، وتبدأ هذه الخطوة بحساب الفرق بين لوغاريتم نسبة الامكان الاعظم للأنموذج في الخطوة السابقة والأنموذج الذي يحتوي على الحد الثابت مع المتغير في الخطوة السابقة و أحد المتغيرات المتبقية. وحسب الصيغة الآتية [6]

$$G_j^2 = \mathrm{MLR}_1 - \mathrm{MLR}_{1+j} \quad (12)$$

إذ أن

MLR_1: لوغاريتم نسبة الامكان للأنموذج في الخطوة السابقة

MLR_{1+j}: لوغاريتم نسبة الامكان للأنموذج الذي يحتوي على الحد الثابت مع المتغير في الخطوة السابقة وأحد المتغيرات المتبقية.

G_j^2: الفرق بين لوغاريتم نسبة الإمكان الأعظم للأنموذجين.

ثم نحسب قيمة (p-value) لكل قيم (G_j^2) ونختار المتغير الذي يحقق الشرط في المعادلة (13) ليكون هو المتغير المرشح للدخول للأنموذج، ثم نختبر الانموذج الجديد باستعمال اختبار الانحراف (D) الذي سبق ذكره , فإذا كان الانموذج المقدر غير معنوي يتم حذف المتغير المضاف ويكون الانموذج الملائم هو الانموذج المقدر في الخطوة السابقة اما اذا كان الانموذج المقدر معنويا فيتم ادراج المتغير الجديد في الانموذج وننتقل الى الخطوة اللاحقة .

وهكذا نستمر بإضافة المتغيرات المستقلة الواحد بعد الأخر، ونتوقف عن الاضافة عندما تكون قيمة (p-value) التي تحقق الشرط في المعادلة (13) اكبر من مستوى المنوية المحدد (α) او عندما يكون الانموذج غير معنوي حسب اختبار الانحراف (D) .

كما يمكن الاعتماد على قيمة (p-value) لأحصاءة والد بدل قيمة (p-value) للفرق بين نسبة الامكان الاعظم في ترشيح المتغير المستقل للدخول الى الانموذج، إذ يتم اختيار المتغير الذي يعطي اقل قيمة (p-value) حسب اختبار والد (wald) للدخول الى الانموذج مع مراعاة ان تكون قيمتها اقل من مستوى المعنوية المحدد (α) ويتم اختبار الانموذج باستعمال اختبار الانحراف (D) [5] وسنعتمد في دراستنا على قيمة (p-value) للفرق بين نسبة الامكان الاعظم في ادخال المتغيرات المستقلة.

طريقة الحذف العكسي (Backward Elimination Method)

على عكس الطريقة السابقة التي تبدأ بافتراض الأنموذج يحتوي على الحد الثابت فقط. فأن هذه الطريقة تبدأ بافتراض ان الانموذج يحتوي على جميع المتغيرات المستقلة، ثم نبدأ بحذف المتغيرات بشكل تدريجي وحسب الخطوات الآتية:

الخطوة الاولى: كما في الطريقة السابقة يتم تطبيق (j) من النماذج وبعدد المتغيرات المستقلة وفي كل انموذج يحذف متغير يختلف عن المتغير السابق ثم نجد الفرق بين قيمتي نسبة الامكان الاعظم للأنموذج بجميع المتغيرات و الانموذج بعد حذف متغير واحد من المتغيرات المستقلة وكما في الصيغة الآتية[6] :

$$G_j^1 = MLR_m - MLR_{m-J} \quad (13)$$

j= 1,2,3.....m

إذ أن

MLR_m: لوغاريتم نسبة الامكان الاعظم للأنموذج بجميع المتغيرات المستقلة.

MLR_{m-j}: لوغاريتم نسبة الامكان الاعظم للأنموذج بجميع المتغيرات عدا المتغير (j).

G_j^1: الفرق في لوغاريتم نسبة الامكان للأنموذجين والتي تتبع توزيع مربع كاي بدرجة حرية تساوي واحد اذا كان المتغير المرشح للحذف كمياً او تساوي عدد صفات المتغير المرشح للحذف مطروحاً منها واحد اذا كان المتغير وصفياً (فئوياً) .ثم نحسب (p-value) لكل قيم (G_j^1) ونختار المتغير الذي يحقق الشرط الآتي للخروج من الانموذج [1] .

$$p_j^{(1)} = \max[p_j^{(1)}] \quad (14)$$

إذ أن

$p_j^{(1)}$: قيمة (p-value) للفرق G_j^1 في الخطوة (1) للمتغير (j) ويجب ان يكون اكبر من مستوى المعنوية المحدد (α).

ثم نختبر الانموذج الذي يتضمن جميع المتغيرات المستقلة عدا المتغير المحذوف حسب اختبار كاي تربع للانحراف (D) الذي تم ذكره، فاذا كان الانموذج غير معنوي يتم التوقف من حذف المتغيرات ويكون الانموذج الملائم هو الانموذج الذي يحتوي على الجميع المتغيرات المستقلة، اما اذا كان الانموذج معنويا فيتم حذف المتغير ثم ننتقل الى الخطوة الآتية.

الخطوة الثانية: بعد ان اصبح لدينا انموذج انحدار لوجستي يحتوي على جميع المتغيرات عدا متغير واحد، فان هذه الخطوة تبدأ بتوفيق (j-1) من النماذج بعد حذف متغير مختلف ومن ثم نحسب الفرق بين لوغاريتم نسبة الامكان الاعظم للأنموذج في الخطوة السابقة و الانموذج الذي يحتوي على جميع المتغيرات المستقلة عدا المتغير الذي تم حذفه في الخطوة السابقة و أحد المتغيرات المتبقية وحسب الصيغة الآتية[6]

$$G_j^2 = MLR_{m-1} - MLR_{m-1-j} \qquad (15)$$

j= 1,2,3.....m-1

إذ أن

MLR_{m-1} : لوغاريتم نسبة الامكان الاعظم للأنموذج في الخطوة السابقة

MLR_{m-1-j} : لوغاريتم نسبة الامكان الاعظم للأنموذج في الخطوة (1) مع حذف المتغير (j)

G_j^2 : الفرق بين لوغاريتم نسبة الامكان الاعظم للأنموذجين

ثم نحسب (p-value) لكل قيم (G_j^2) ونختار المتغير الذي يحقق الشرط في المعادلة (14) ليكون هو المتغير المرشح للخروج من الانموذج، ثم نختبر الانموذج بشكل عام حسب اختبار مربع كاي للانحراف (D) , فاذا كان الانموذج غير معنوي يتم التوقف عن حذف المتغيرات ويكون الانموذج الملائم هو الانموذج الذي تم التوصل اليه في الخطوة الاولى , اما اذا كان الانموذج معنوياً فيتم حذف المتغير ثم ننتقل الى الخطوة اللاحقة [6]

وهكذا نستمر بحذف المتغيرات المستقلة الواحد بعد الأخر، ونتوقف عن الحذف عندما تكون قيمة (p-value) التي تحقق الشرط في المعادلة (14) اقل من مستوى المعنوية المحدد (α)، او عندما يكون الانموذج غير معنوي حسب اختبار الانحراف (D).

كما يمكن الاعتماد على قيمة (p-value) لأحصاءة والد بدل قيمة (p-value) للفرق بين نسبة الامكان الاعظم في ترشيح المتغير المستقل للحذف من الانموذج، إذ يتم اختيار المتغير الذي يعطي أكبر قيمة (p-value) للحذف من الانموذج مع مراعاة ان تكون قيمتها أكبر من مستوى المعنوية المحدد (α) ويتم اختبار الانموذج باستعمال اختبار الانحراف (D) [5] وسنعتمد في دراستنا على قيمة (p-value) للفرق بين نسبة الامكان الاعظم في حذف المتغيرات المستقلة.

معايير اختيار أفضل انموذج انحدار لوجستي

معيار نسبة الامكان الاعظم Maximum Likelihood Ratio (MLR)

يعتمد هذا المعيار في تقييمه لأداء الانموذج على امكانية حدوث نتائج البيانات التي تم اعتمادها اذا ما عرفنا تقديراً لمعلمات الانموذج مع اختيار تقديراً للمعلمات التي تعظم امكانية الحصول على النتائج المشاهدة (المعالم المقدرة

بطريقة الامكان الاعظم) وبما ان قيمة الامكان الاعظم دائما تكون صغيرة جداً وهي اقل من الواحد , لذلك نأخذ لها اللوغاريتم الطبيعي ونضربها (-2) للتخلص من القيمة السالبة الناتجة من لوغاريتم دالة الامكان الاعظم , وبهذا يكون معيار المفاضلة هو (-2) مضروب في لوغاريتم الامكان الاعظم والمسمى بنسبة الامكان الاعظم ويمكن حسابه كما في المعادلة الآتية

$$MLR = -2\ln\prod_{i=1}^{n}\hat{p}_i^{yi}\,(1-\hat{p}_i)^{1-yi} \qquad (16)$$

إذ أن

$\hat{p}_i$ هي نسبة الاستجابة المتوقعة لكل مشاهدة من مشاهدات المتغير المعتمد [6].

وكلما كانت قيمة دالة الامكان كبيرة تكون قيمة معيار نسبة الامكان الاعظم صغيرة يدل هذا على جودة الأنموذج المقدر اي ان كلما كانت قيمة معيار نسبة الامكان الاعظم اقل كان الانموذج افضل [5].

معيار جدول التصنيف Classification Table Criterion (CTC)

يوضح جدول التصنيف عدد الحالات المشاهدة التي تمتلك صفة معينة وعدد الحالات المشاهدة التي لا تمتلك هذه الصفة مقابل الحالات المتوقعة والتي صنفها الانموذج بأنها تمتلك تلك الصفة والحالات المتوقعة التي لا تمتلك تلك الصفة بحيث يبين الجدول نسبة ملاءمة الحالات المشاهدة للحالات المتوقعة [2] , ومن المعروف أن قيم المشاهدات المتوقعة هي قيم احتمالية محصورة قيمها بين الواحد الصحيح والصفر ولمعرفة عدد الحالات المتوقعة التي تصنف بأنها تمتلك تلك الصفة من الحالات المتوقعة التي لا تمتلكها فيجب أن تكون هناك نقطة معينة تفصل بين هاتين الحالتين وهذه النقطة تسمى بنقطة القطع وهي في الاغلب تساوي (0.5) فالحالات التي يكون ناتج احتمالها المتوقع أكبر من أو يساوي نقطة القطع (0.5) تصنف بأنها حالات حامله لتلك الصفة والحالات التي يكون احتمالها اقل من نقطة القطع (0.5) تصنف بأنها حالات لا تحمل تلك الصفة [5] وتتم مقارنتها مع التكرار المشاهد وكما في الجدول الآتي :

جدول (1)

الشكل العام لجدول التصنيف

المشاهد \ المتوقع	y=0	y=1	نسبة التصنيف الصحيح
y=0	NN	NT	$\frac{NN}{NN+NT}$
y=1	TN	TT	$\frac{TT}{TN+TT}$

إذ أن

NN : الحالة المشاهدة تساوي صفراً و المتوقعة اقل من (0.5)

NT : الحالة المشاهدة تساوي صفراً والمتوقعة اكبر من او يساوي (0.5)

TN : الحالة المشاهدة تساوي واحداً والمتوقعة اقل من (0.5)

TT : الحالة المشاهدة تساوي واحداً والمتوقعة اكبر من او يساوي (*0.5*)

$\frac{NN}{NN+NT}$: نسبة التصنيف الصحيح للحالة المشاهدة تساوي صفراً ويسمى بالحساسية

$\frac{TT}{TN+TT}$: نسبة التصنيف الصحيح للحالة المشاهدة تساوي واحداً ويسمى بالدقة

ويمكن ان نحسب نسبة التصنيف الصحيح للأنموذج ككل من الصيغة الآتية:

$$\text{نسبة التصنيف الصحيح للأنموذج} = \frac{NN+TT}{\text{مجموع الحالات الكلي}} * 100 \qquad (17)$$

وكلما كانت النسبة مرتفعة دل ذلك على جودة الانموذج المقدر.

بعد التطرق الى مفهوم الانحدار اللوجستي والاختبارات التي تتعلق بجودة التوفيق للأنموذج والمتغيرات داخل الانموذج وطرائق اختيار انموذج الانحدار اللوجستي ومعايير اختيار الانموذج وسيتم تطبيق انموذج الانحدار اللوجستي على بيانات المرضى الراقدين في المستشفيات بوباء كورونا، وسيتم استعمال طريقتي التقدم الامامي والحذف العكسي لتحديد اهم العوامل المؤثرة في وفيات وباء كورونا , وتتم المقارنة بين نتائجهما عن طريق معايير اختيار افضل انموذج. وكما مبين في الفقرة القادمة.

الجانب العملي

سيتم تطبيق انموذج الانحدار اللوجستي على المرضى الراقدين في مستشفى سوق الشيوخ لسنة 2020 الذين بلغ عددهم (439) كان منهم (43) فارقوا الحياة و (398) تماثلوا للشفاء لمعرفة العوامل التي لها اكثر تأثير في وفيات هؤلاء المصابين , وأخذت المعلومات من طبلة المريض الراقد ومن سجل المرضى الراقدين فضلاً عن برنامج الراقدين المصمم من لدن وزارة الصحة العراقية.

وصف المتغيرات

تم استعمال تسعة متغيرات منها متغير واحد تابع (y) ثنائي القيمة يأخذ صفرا لحدوث الوفاة والواحد لحالات الشفاء وثمانية متغيرات مستقلة سيتم تعريفها كما في الجدول الاتي :

جدول (2)

تعريف المتغيرات المستقلة

رمز المتغير	تمثيل المتغير	القيم التي يأخذها	تمثيل كل قيمة	رمز المتغير	تمثيل المتغير	القيم التي يأخذها	تمثيل كل قيمة
X_1	متغير وصفي يمثل جنس المريض الراقد	1	ذكر	X_3	متغير وصفي يمثل مهنة المريض الراقد	1	طفل
		2	انثى			2	طالب
X_2	متغير وصفي يمثل التحصيل الدراسي للمريض الراقد	1	امي			3	موظف
		2	ابتدائية			4	اعمال حرة
		3	متوسطة			5	ربة بيت
		4	اعدادية			6	متقاعد
		5	دبلوم وبكالوريوس			7	عاجز
		6	ماجستير ودكتوراه	X_6	متغير وصفي يمثل محل السكن للمريض الراقد	1	حضر
X_4	متغير وصفي يمثل الحالة الاجتماعية	1	أعزب			2	ريف
		2	متزوج	X_7	متغير كمي يمثل فترة رقود المريض بالأيام		
		3	**أرمل او ارملة او مطلق او مطلقة**	X_8	متغير وصفي يمثل التدخين	1	مدخن
X_5	متغير كمي يمثل عمر المريض بالسنوات					2	غير مدخن

النتائج

تم استعمال البرنامج الاحصائي (SPSS 24) وانموذج الانحدار اللوجستي حسب المعادلة الآتية:

$$\ln\left(\frac{p}{q}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots\dots\dots + \beta_m X_m \qquad (18)$$

وكانت النتائج التي تم الحصول عليها والتي ستتم مناقشتها مع تبويبها في جداول ليسهل تحليلها وفهمها. اذ أظهرت النتائج ان قيمة نسبة الامكان الاعظم مع الحد الثابت فقط عند الخطوة التكرارية السابعة هي (966.909). وسنعرض نتائج طريقتي الاختيار مع المقارنة بينها عن طريق معايير اختيار أفضل انموذج وسنستعمل مستوى ثقة مقداره (95%) لترشيح المتغير للدخول الى الانموذج ومستوى ثقة مقداره (99%) لتثبيت المتغير في الانموذج

اولا: طريقة التقدم الامامي

بدأت هذه الطريقة بافتراض ان الانموذج يحتوي على الحد الثابت فقط ثم تم ترشيح المتغير المستقل (X_8) وهو التدخين في الخطوة (1) الذي له اصغر (p-value) و تساوي (10^{-77}) لقيمة الفرق بين نسبة الامكان الاعظم للأنموذج مع الحد الثابت و الانموذج مع الحد الثابت والمتغير (X_8) التي تساوي (G_8^1=346.461) والتي تم حسابها على

وفق الصيغة (10) وهي اقل من مستوى المعنوية المحدد (α=0.05) و لتثبيت المتغير(X_8) في الانموذج نستعمل اختبار الانحراف (D) الذي قيمته تساوي قيمة (G_8^1) ومن ثم يعطي قيمة (p-value) نفسها وهي اقل من مستوى المعنوية المحدد (α=0.01) لذا سيتم تثبيت المتغير(X_8) في الانموذج , ثم ننتقل الى الخطوة (2), التي نتج عنها ترشيح المتغير (X_5) وهو عمر المريض , ولتثبيته في الانموذج نستعمل اختبار الانحراف (D) الذي قيمته (360.017) وكانت قيمة (p-value= 0.000) وهي اقل من مستوى المعنوية المحدد (α= 0.01) لهذا تم تثبيت المتغير (X_5) بالأنموذج , وبعد ان تم حساب قيم (p-value) لـ (G_j^2) إذ $j = 1,2, \dots .7 \quad j \neq 5$ تبين ان جميعها اكبر من مستوى المعنوية (α=0.05) لذا تم التوقف عن اضافة المتغيرات, وأصبح الانموذج النهائي هو الانموذج الناتج من الخطوة الثانية الذي يحتوي على الحد الثابت مع المتغيرات المستقلة الآتية (X_8, X_5) , وكانت نتائج المعلمات المقدرة كما مبينة في المعادلة الاتية:

$$\ln\left(\frac{p}{1-p}\right) = -14.64 + 0.109X_5 + 4.38X_{8(1)} \qquad (19)$$

ولاختبار معنوية الانموذج المقدر في المعادلة (19) بشكل عام تم استعمال اختبار هوزمر – ليمشو (H&L) والذي كانت قيمته التي تم حسابها على وفق الصيغة (7) تساوي (0.929) و ان قيمة (p-value) له تساوي (0.996) وهي اكبر من مستوى المعنوية (α=0.05) وهذا يعني عدم رفض فرضية العدم (6), وان الانموذج المقدر يمثل البيانات بشكل جيد .

ثانيا: طريقة الحذف العكسي

بدأت هذه الطريقة بافتراض ان الانموذج يحتوي على جميع المتغيرات المستقلة ثم تبدأ بحذف المتغير المستقل الذي له اكبر قمة (p-value) للفرق بين نسبة الامكان الاعظم للأنموذج بجميع المتغيرات المستقلة ونسبة الامكان الاعظم لجميع المتغيرات عدا متغير واحد وبدرجة حرية تساوي عدد تكرارات المتغير مطروحاً منها واحد وبعد ان تم حساب قيم (p-value) لجميع المتغيرات كان المتغير (X_7) مدة رقود المريض هو المتغير الذي له اكبر قيمة (p-value) والتي تساوي (0.82) لذا تم ترشيحه للحذف من الانموذج , ولتأكيد حذف المتغير(X_7) نستعمل اختبار الانحراف (D) الذي تم حسابه على وفق الصيغة (5) وكانت قيمته (599.121) وقيمة (p-value) له هي (0.000) وهي اقل من مستوى المعنوية (α=0.01) لذا تم حذف المتغير بشكل نهائي من الانموذج , وهكذا نستمر بحذف المتغيرات ويتم التوقف عن الحذف عندما تكون اكبر قيم (p-value) لجميع المتغيرات المتبقية اقل من مستوى المعنوية المحدد (α=0.05) او يكون الانموذج غير معنوي حسب اختبار الانحراف, وهو ما تحقق عنده الخطوة السابعة إذ أن اكبر قيم (p-value) هي (0.039) التابعة للمتغير (X_5) عمر المريض هي اقل من (0.05) لذلك لم يتم حذفه من الانموذج وبهذا كان الانموذج المقدر هو الانموذج الناتج عن الخطوة السادسة و يمكن كتابة الانموذج الناتج كما في المعادلة الآتية :

$$\ln\left(\frac{p}{1-p}\right) = \quad -14.64 + 0.109X_5 + 4.38X_{8(1)} \qquad (20)$$

وهو الانموذج نفسه الذي نتج من طريقة التقدم الامامي وبينت نتائج اختبار (H&L) ان الانموذج المقدر معنوي إذ كانت قيمة ($x^2 = 0.929$) الذي تم حسابه على وفق الصيغة (7) وكانت قيمه (p-value) له تساوي (0.996) وهي أكبر من مستوى المعنوية المحدد (α=0.05) وهذا يعني عدم رفض فرضية العدم (6) والانموذج المقدر يمثل البيانات بشكل جيد.

جدول (3)

اختبارهوزمر- ليمشو لطريقة التقدم الامامي عند الخطوة الثانية

Step	Chi-square	df	p.value
2	.929	7	.996

جدول (4)

اختبارهوزمر- ليمشو لطريقة الحذف العكسي عند الخطوة السابعة

Step	Chi-square	Df	p.value
7	.929	7	.996

المقارنة بين طريقتي التقدم الامامي والحذف العكسي

كل طريقة من الطريقتين السابقتين افرزت لنا انموذج انحدار لوجستي بمتغيرين هما عمر المريض و التدخين وان معايير المفاضلة بين الطريقتين كانت متساوية ونبدأ من المعيار الاول وهو نسبة الامكان الاعظم (MLR) إذ نلاحظ ان قيمته (132.475) التي تم حسابها على وفق المعادلة (16)، اما المعيار الثاني وهو جدول التصنيف فكانت نسبة التصنيف الصحيح (95.7%) والتي تم حسابها على وفق المعادلة (17)وكما مبين بالجدول التالي،

جدول (5)

Observed		Predicted		
		Y		Percentage Correct
		0	1	
y	0	392	4	99.0
	1	15	28	65.1
Overall Percentage				95.7

الاستنتاجات والتوصيات

من تطبيق طريقتي اختيار أفضل انموذج انحدار لوجستي ثنائي، يمكن الإشارة الى بعض الاستنتاجات التي توصل اليها الباحثون والتوصيات التي يعتقد انها ضرورية.

الاستنتاجات

1. ان جميع نماذج الانحدار اللوجستي النهائية الناتجة من طريقتي اختيار أفضل انموذج انحدار لوجستي قد اجتازت اختباري جودة التوفيق للأنموذج بشكل عام وهما اختبار الانحراف (D) واختبار هوزمر – ليمشو (H&L).
2. عند استعمال طريقة التقدم الامامي او طريقة الحذف العكسي فان المتغيرات التي لها اكثر تأثير في وفيات المرضى الراقدين بوباء كورونا هي (عمر المريض، التدخين)
3. ومن نتائج الجانب العملي وعند استعمال طريقتي التقدم الامامي والحذف العكسي تبين ان المتغير التدخين هو المتغير الاكثر تأثيراً في وفيات المرضى المصابين.

التوصيات

استناداً الى ما جاء بالجوانب النظرية والتجريبية والتطبيقية وعلة وفق الاستنتاجات التي تم التوصل اليها نوصي بما يلي:

1. ضرورة اختيار المتغيرات المستقلة بعناية تامة من قبل الباحث حتى يتم الحصول على النتائج بصورة دقيقة.
2. توعية المجتمع وخاصة المصاب منهم بوباء كورونا بعدم التدخين لما له من تأثير قوي على حياتهم.
3. ضرورة تشجيع و توسيع البحوث التي تخص الجانب الصحي وخاصة البحوث التي تتعلق بالأمراض والاوبئة التي لها تأثير في حياة الناس بصوره مباشرة.

المصادر

اولاً: المصادر العربية

1. **ابراهيم، جمال محمد شاكر محمد (1997) "** دراسة احصائية للعوامل المسببة لمرض سرطان الدم وتلك المسببة للتمايز بين المرضى بعد مرحلة العلاج بالتطبيق على مستشفيات جامعة المنصورة " رسالة ماجستير، جامعة المنصورة ,كلية التجارة , مصر
2. **بابطين، عادل احمد حسن (2009)** "الانحدار اللوجستي وكيفية استعماله في بناء نماذج التنبؤ للبيانات ذات المتغيرات التابعة ثنائي القيمة " اطروحة دكتوراه، جامعة ام القرى , كلية التربية قسم علم النفس , الرياض , المملكة العربية السعودية.
3. **عباس، علي خضر (2012)** "استعمال نموذج الانحدار اللوجستي في التنبؤ بالدوال ذات المتغيرات التابعة النوعية " مجلة جامعة كركوك للعلوم الادارية والاقتصادية ,العراق , المجلد 2 , العدد 2
4. **الفرهود , سهيلة حمود عبد الله (2014)** "استعمال الانحدار اللوجستي لدراسة العوامل المؤثرة على اداء الاسهم " joural of Al Azhar university-Gaze 2014,16:47-68 .
5. **فهمي , محمد شامل بهاء الدين ,(2005)** "الاحصاء بلا معاناة :المفاهيم مع التطبيقات على برنامج spss" الادارة العامة للباعة والنشر , معهد الادارة العامة , الرياض , المملكة العربية السعودية .

ثانياً : المصادر الاجنبية

6. **Hosmer , David .W , and Lemeshow, S (2000)** "Applied Logistic Regression" 2nd, edition ,John wiley and sons , New York .
7. **Kleinbaum , David & Klein , Mitchel (2002) "** Logistic Regression" a self – leavning TeeXt , USA . springer .
8. **Pample, Fred C. (2000).** Logistic Regression Aprimer. Sage University Paper series on Quantitative Applications in the Social Sciencies. No.07-132). Beverly Halls, CA: Sage.
9. **Walker, Jonathan (1996)** "Methodology Application: Logistic Regression Using the codes Data" Developed For Department of Transportation, National Highway Traffic Safety Administration (NHTS), Washington DC, April 30, 1996 .